4F7: STATISTICAL SIGNAL ANALYSIS

S.S. Singh, Easter 2021

Question 1. Part (a-i)

Write in vector form to get $\widehat{X}_1 - X_1 = h^T(Y - m) - (X_1 - b_1)$. Square both sides and take the expectation to get

$$h^T \Sigma_{y,y} h + (\Sigma_{x,x})_{1,1} - 2h^T (\Sigma_{y,x})_{:,1}$$

where $(\Sigma_{y,x})_{:,1}$ denotes the first column of the matrix. Differentiate and set to 0 to get

$$\Sigma_{y,y} h = (\Sigma_{y,x})_{:,1}$$

or

$$h = (\Sigma_{y,y})^{-1} (\Sigma_{y,x})_{:,1}$$

Part (a-ii)

Substitute the h from previous part to get

$$\begin{aligned}
&h^T \Sigma_{y,y} h + (\Sigma_{x,x})_{1,1} - 2h^T (\Sigma_{y,x})_{:,1} \\
&= h^T (\Sigma_{y,x})_{:,1} + (\Sigma_{x,x})_{1,1} - 2h^T (\Sigma_{y,x})_{:,1} \\
&= (\Sigma_{x,x})_{1,1} - h^T (\Sigma_{y,x})_{:,1} \\
&= (\Sigma_{x,x})_{1,1} - (\Sigma_{y,x})_{:,1}^T (\Sigma_{y,y})^{-1} (\Sigma_{y,x})_{:,1}
\end{aligned}$$

Part (a-iii)

From the formulation of the previous parts a significant simplification is noted. Each of the q-terms of the objective function can be solved separately to get the minimising row i of matrix H is the vector

$$h = \left[(\Sigma_{y,y})^{-1} (\Sigma_{y,x})_{:,i}\right]^T = \left[(\Sigma_{y,x})_{:,i}\right]^T (\Sigma_{y,y})^{-1} = (\Sigma_{x,y})_{i,:} (\Sigma_{y,y})^{-1}$$

Thus $H = (\Sigma_{x,y}) (\Sigma_{y,y})^{-1}$.

Part (b-i)

$$X_k^2 = a^2 X_{k-1} + W_k^2 + 2aX_{k-1}W_k.$$

Noting that $\mathbb{E}(X_{k-1}W_k) = 0$, we get $\sigma_0^2 = a^2\sigma_0^2 + \sigma_w^2$ and solve for a to get

$$a^2 = 1 - \frac{\sigma_w^2}{\sigma_0^2}.$$

Part (b-ii)

To find $\Sigma_{x,x}$, find the autocorrelation of X_k.

$$\mathbb{E}(X_k X_{k+m}) = \mathbb{E}(X_k \left[W_{k+m} + aW_{k+m-1} + \cdots + a^{m-1}W_{k+1} + a^m X_k\right])$$

which evaluates to $a^m\sigma_0^2$. Now assemble the entries of Σ_{xx} to get (noting that the matrix is symmetric)

$$\begin{bmatrix} \sigma_0^2 & a\sigma_0^2 & & \cdots & a^{p-1}\sigma_0^2 \\ a\sigma_0^2 & \sigma_0^2 & a\sigma_0^2 & & a^{p-2}\sigma_0^2 \\ & & \ddots & & \\ & & & & \vdots \\ & & & & \sigma_0^2 \end{bmatrix}$$

$[Y_1, \ldots, Y_q] = [X_1, \ldots, X_q] + [V_1, \ldots, V_q]$. Thus

$$\Sigma_{x,y} = \mathbb{E}\left(\begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} [X_1, \ldots, X_q]\right)$$

which is the $p \times q$ submatrix of $\Sigma_{x,x}$ assuming $p \geq q$. (Marks will also awarded if it is assumed $p \leq q$.)

Similarly $\mathbf{Y} = \mathbf{X} + \mathbf{V}$ and thus $\Sigma_{y,y}$ is found as follows

$$\mathbb{E}\left(\mathbf{Y}\mathbf{Y}^T\right) = \mathbb{E}\left(\mathbf{X}\mathbf{X}^T\right) + \mathbb{E}\left(\mathbf{V}\mathbf{V}^T\right) + \mathbb{E}\left(\mathbf{X}\mathbf{V}^T\right) + \mathbb{E}\left(\mathbf{V}\mathbf{X}^T\right)$$

Cross terms have zero expectation. $\mathbb{E}\left(\mathbf{V}\mathbf{V}^T\right)$ is a $q \times q$ diagonal matrix with entries σ_v^2. $\mathbb{E}\left(\mathbf{X}\mathbf{X}^T\right)$ is the submatrix of $\Sigma_{x,x}$ using the first q rows and columns. Thus $\Sigma_{y,y} = \left(\Sigma_{x,x}\right)_q + \sigma_v^2 I$.

Question 2. Part (a)

$v_1 = v_1(1-a) + v_2 b$ or $av_1 = bv_2$.
$v_2 = v_1 a + v_2(1-a-b) + v_3 b$ or $av_2 = bv_3$. Check $av_k = bv_{k+1}$
$v_k = (a/b)^{k-1} v_1$
Sum by geometric series to solve for v_1

$$\begin{aligned}\sum_{k=1}^{\infty} v_k &= 1 \\ \sum_{k=1}^{\infty} v_1 \left(\frac{a}{b}\right)^{k-1} &= 1 \\ v_1 &= 1 - \frac{a}{b}.\end{aligned}$$

Part (b-i)
The update step:

$$\bar{\pi}_n(i) = \frac{\pi_n(i)\lambda i \exp(-\lambda i y_n)}{\sum_{j=1}^{\infty} \pi_n(j)\lambda j \exp(-\lambda j y_n)}$$

The prediction step:

$$\pi_{n+1}(j) = \sum_{i=1}^{\infty} \bar{\pi}_n(i) p_{i,j}$$

which gives $\pi_{n+1}(j) = \bar{\pi}_n(j-1)a + \bar{\pi}_n(j)(1-a-b) + \bar{\pi}_n(j+1)b$ for $j > 1$ and $\pi_{n+1}(1) = \bar{\pi}_n(1)(1-a) + \bar{\pi}_n(2)b$.
Part (b-ii)

$$\begin{aligned}\beta_n(i) &= p(y_{n:T}|X_n = i) \\ &= \sum_j p(y_{n:T}, X_{n+1} = j|X_n = i) \\ &= \sum_j p(y_{n:T}|X_{n+1} = j, X_n = i)p(X_{n+1} = j|X_n = i) \\ &= \sum_j p(y_n|X_n = i)p(y_{n+1:T}|X_{n+1} = j)p_{i,j} \\ &= \sum_{j=i-1}^{i+1} \lambda i \exp(-\lambda i y_n)\beta_{n+1}(j)p_{i,j}\end{aligned}$$

Part (b-iii)

$$\Pr(X_n = i|y_{1:T}) = \frac{\Pr(X_n = i|y_{1:n-1})p(y_{n:T}|X_n = i)}{\sum_{j=1}^{\infty} \Pr(X_n = j|y_{1:n-1})p(y_{n:T}|X_n = j)}$$
$$= \frac{\pi_n(i)\beta_n(i)}{\sum_{j=1}^{\infty} \pi_n(j)\beta_n(j)}.$$

Marks will be awarded for a more detailed derivation of the first equality.

Part (c)

$$p(x_1, y_1, \ldots, x_T, y_T|x_0) = p(x_{1:T}|x_0) \prod_{n=1}^{T} \lambda x_n \exp(-\lambda x_n y_n)$$

Take the log of the expression

$$\log p(x_{1:T}, y_{1:T}|x_0) = \log p(x_{1:T}|x_0) + T \log \lambda - \lambda \sum_{n=1}^{T} x_n y_n + C$$

where C is a non-λ term. Differentiate with respect to λ and set to zero

$$\frac{1}{\lambda} = \frac{1}{T} \sum_{n=1}^{T} x_n y_n$$

Part (d)

The Expectation-Maximisation (EM) algorithm is comprised of the E-step and the M-step. Let $\widehat{\lambda}$ denote the current best guess of parameter λ. Compute $p(x_n|y_{1:T})$, for $n = 1, \ldots, T$ with the current estimate $\widehat{\lambda}$ and then compute the Q-function

$$\begin{aligned} Q(\lambda) &= \sum_{i=1}^{T} \sum_{x_i=1}^{\infty} p(x_i|y_{1:T}) \log p(y_i|x_i) \\ &= \sum_{i=1}^{T} \sum_{x_i=1}^{\infty} p(x_i|y_{1:T}) \left(\log \lambda - \lambda x_i y_i + \log x_i\right) \\ &= C + T \log \lambda + \sum_{i=1}^{T} \sum_{x_i=1}^{\infty} p(x_i|y_{1:T}) \left(-\lambda x_i y_i\right) \\ &= C + T \log \lambda + \sum_{i=1}^{T} \mathbb{E}\left\{X_i|y_{1:T}\right\} \left(-\lambda y_i\right) \\ &= C + T \log \lambda - \lambda S \end{aligned}$$

where C is a non-λ term and

$$S = \sum_{i=1}^{T} \mathbb{E}\left\{X_i|y_{1:T}\right\} y_i.$$

The the M-step then maximises $Q(\lambda)$ to get the new best estimate of λ. This is done by differentiating to get

$$\lambda = T/S.$$

This E-M steps are repeated until convergence of the estimate is observed.

Question 3. Part (a)

Note that $\mathbb{E}(X_n) = \mu_0$ for all n. For unbiased estimation need $\mathbb{E}(\widehat{X}_n) = \mathbb{E}(X_n) = \mu_0$ or

$$\begin{aligned}\mu_0 &= L_n \mathbb{E}(\widehat{X}_{n-1}) + K_n \mathbb{E}(Y_n) \\ &= L_n \mu_0 + K_n \mu_0\end{aligned}$$

which implies $L_n + K_n = 1$.

Part (b)-(i)

Let $e_n = X_n - \widehat{X}_n$.

$$\begin{aligned}\widehat{X}_n &= \widehat{X}_{n-1} + G_n(Y_n - \widehat{X}_{n-1}) \\ e_n &= X_n - \widehat{X}_{n-1} - G_n(Y_n - \widehat{X}_{n-1}) \\ &= e_{n-1} + W_n - G_n(X_{n-1} + W_n + V_n - \widehat{X}_{n-1}) \\ &= e_{n-1}(1 - G_n) + W_n(1 - G_n) - G_n V_n\end{aligned}$$

Part (b)-(ii)

Square it and take the expectation to get (noting cross terms have zero expectation):

$$\mathbb{E}(e_n^2) = \mathbb{E}(e_{n-1}^2)(1 - G_n)^2 + \sigma_w^2(1 - G_n)^2 + G_n^2 \sigma_v^2$$

Differentiating the right-hand side with respect to G_n and equating to 0 to solve for G_n yields:

$$G_n = \frac{\sigma_{n-1}^2 + \sigma_w^2}{\sigma_{n-1}^2 + \sigma_w^2 + \sigma_v^2}.$$

When $\sigma_v^2 = 0$ then $Y_n = X_n$ and indeed $G_n = 1$ as required.

Part (c-i)

Using a previous part where it was found that

$$\mathbb{E}(e_n^2) = \mathbb{E}(e_{n-1}^2)(1 - G_n)^2 + \sigma_w^2(1 - G_n)^2 + G_n^2 \sigma_v^2.$$

Substitute $G_n = \sigma_{n-1}^2/(\sigma_{n-1}^2 + \sigma_v^2)$ which gives

$$\begin{aligned}
\sigma_n^2 &= \sigma_{n-1}^2 \left(\frac{\sigma_v^2}{\sigma_{n-1}^2 + \sigma_v^2} \right)^2 + \left(\frac{\sigma_{n-1}^2}{\sigma_{n-1}^2 + \sigma_v^2} \right)^2 \sigma_v^2 \\
\frac{\sigma_n^2}{\sigma_v^2} &= \sigma_{n-1}^2 \sigma_v^2 \left(\frac{1}{\sigma_{n-1}^2 + \sigma_v^2} \right)^2 + \left(\frac{\sigma_{n-1}^2}{\sigma_{n-1}^2 + \sigma_v^2} \right)^2 \\
&= H_n^2 \left(\frac{\sigma_v^2}{\sigma_{n-1}^2} + 1 \right) \\
&= \frac{\sigma_{n-1}^2}{\sigma_{n-1}^2 + \sigma_v^2} \\
&= H_n
\end{aligned}$$

Part (c.ii)

The mean square error of the sample mean estimate is

$$\mathbb{E}\left\{ \left(X_n - \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^2 \right\} = \mathbb{E}\left\{ \left(\frac{1}{n} \sum_{i=1}^{n} V_i \right)^2 \right\} = \frac{\sigma_v^2}{n}.$$

Part (c.iii)

From a previous part we have

$$\frac{\sigma_n^2}{\sigma_v^2} = \frac{\sigma_{n-1}^2/\sigma_v^2}{\sigma_{n-1}^2/\sigma_v^2 + 1}$$

Using $\sigma_0^2/\sigma_v^2 = 1$, we see that $\sigma_1^2/\sigma_v^2 = 1/2$ and thus$\sigma_2^2/\sigma_v^2 = 1/3$. Assuming $\sigma_{n-1}^2/\sigma_v^2 = 1/n$ and thus

$$\sigma_n^2/\sigma_v^2 = 1/(n+1).$$

This is slightly better than the mean square error of the sample mean estimate which was found to have $\sigma_n^2/\sigma_v^2 = 1/n$.

Part (d)

When $\sigma_w^2 = 0$, $G_n = H_n$, i.e. H_n is the best gain for a static signal.

Question 4. Part (a)
$p(x_k|x_{k-1})$ is $\mathcal{N}(x_k; \alpha x_{k-1}, 1)$. The joint density is

$$\begin{aligned}
\log p(x_{1:k}) &= \log\left(p(x_1|x_0)\cdots p(x_k|x_{k-1})\right) \\
&= \log p(x_1|x_0) + \cdots + \log p(x_k|x_{k-1}) \\
&= -\frac{1}{2}\sum_{i=1}^{k}(x_i - \alpha x_{i-1})^2 + C \\
\frac{\mathrm{d}}{\mathrm{d}\alpha}\log p(x_{1:k}) &= \sum_{i=1}^{k}(x_i - \alpha x_{i-1})x_{i-1}
\end{aligned}$$

where C is the sum of the non-α terms. Differentiate and set to zero to solve for maximising α which is

$$\frac{\sum_{i=1}^{k} x_i x_{i-1}}{\sum_{i=1}^{k} x_{i-1}^2}.$$

Part (b-i)

Using the cdf method, $\Pr(Y_k \le y_k | X_k = x_k) = \Pr(V_k \le y_k - |x_k|)$. Differentiating the cdf gives the pdf which is $p(y_k|x_k) = f(y_k - |x_k|)$ where f is the pdf of V_k.

Part (b-ii)

The prediction step is

$$p(x_{k+1}|y_{1:k}) = \int \mathcal{N}(x_k; \alpha x_{k-1}, 1) p(x_k|y_{1:k})\mathrm{d}x_k$$

The update step is

$$p(x_{k+1}|y_{1:k+1}) = \frac{f(y_{k+1} - |x_{k+1}|)p(x_{k+1}|y_{1:k})}{\int f(y_{k+1} - |x_{k+1}|)p(x_{k+1}|y_{1:k})\mathrm{d}x_{k+1}}.$$

Part (c)

$$p(y_{1:k}) = \int p(y_{1:k}|x_{1:k})p(x_{1:k})\mathrm{d}x_{1:k}$$

Take the derivative of $\log p(y_{1:k})$ to get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha}\log p(y_{1:k}) &= \frac{1}{p(y_{1:k})}\frac{\mathrm{d}}{\mathrm{d}\alpha}p(y_{1:k}) \\
&= \frac{1}{p(y_{1:k})}\frac{\mathrm{d}}{\mathrm{d}\alpha}\int p(y_{1:k}|x_{1:k})p(x_{1:k})\mathrm{d}x_{1:k} \\
&= \frac{1}{p(y_{1:k})}\int p(y_{1:k}|x_{1:k})\frac{\mathrm{d}}{\mathrm{d}\alpha}p(x_{1:k})\mathrm{d}x_{1:k} \\
&= \frac{1}{p(y_{1:k})}\int p(y_{1:k}|x_{1:k})p(x_{1:k})\frac{\mathrm{d}}{\mathrm{d}\alpha}\log p(x_{1:k})\mathrm{d}x_{1:k} \\
&= \int p(x_{1:k}|y_{1:k})\frac{\mathrm{d}}{\mathrm{d}\alpha}\log p(x_{1:k})\mathrm{d}x_{1:k}
\end{aligned}$$

and thus

$$h(x_{1:k}) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\log p(x_{1:k}).$$

The importance sampling estimate is

$$\frac{\sum_{i=1}^{N} h(X_{1:k}^i)w_k^i}{\sum_{i=1}^{N} w_k^i} = 0$$

with the specific form for h from previous parts substituted in.

Part (d-i)

Calculate $\mathbb{E}(h(X_{1:k}^{J_1}))$ to verify result:

$$\begin{aligned}
\mathbb{E}\left[\left(N^{-1}\sum_{i=1}^{N} w_k^i\right) h(X_{1:k}^{J_1})\right] &= \mathbb{E}\left[\left(N^{-1}\sum_{i=1}^{N} w_k^i\right)\sum_{j=1}^{N} h(X_{1:k}^j)\Pr(J_1 = j)\right] \\
&= \mathbb{E}\left[\left(N^{-1}\sum_{i=1}^{N} w_k^i\right)\sum_{j=1}^{N} h(X_{1:k}^j)\frac{w_k^j}{\sum_{i=1} w_k^i}\right] \\
&= \mathbb{E}\left[N^{-1}\sum_{j=1}^{N} h(X_{1:k}^j)w_k^j\right]
\end{aligned}$$

and the final term has the required mean.

Part (d-ii)

The importance sampling estimate with resampling is

$$\frac{\sum_{i=1}^{N} g(X_{1:k+1}^i)w_{k+1}^i}{\sum_{i=1}^{N} w_{k+1}^i}$$

where

$$w_{k+1}^i = \left(\frac{1}{N}\sum_{j=1}^{N} w_k^j\right) \times p(y_{k+1} \mid X_{k+1}^i)$$

and X_{k+1}^i is a sample from $p(x_{k+1} \mid X_k^{J_i}) = \mathcal{N}(x_{k+1}|\alpha X_k^{J_i}, 1)$.

Examiner's comments:

Q1: A popular question answered well by most. Some failed to realise that their solution from part (a)-(i) could be used to inform their answer for part (a)-(iii). There were some supprising outcomes like a small number of candidates not being able to calculate the autocorrelation function for an AR(1) process needed in part (b)-(ii).

Q2: A popular question with many doing very well in parts (b) and (c). Part (a) was not answered in full by a significant number eventhough they had executed the equation for v correctly.

Q3: This was a very well answered question although some did not know the definition of an unbiased estimate, which is a Part IIA concept. Clearly the majority of the students have mastered the relevant concepts well.

Q4: This was a well answered question in the main. Many were able to successfully employ importance sampling for MLE estimate in part (c). The proof in part (d) was also executed well. It was a surprise to see a small number struggle with the elementary concept of part (b)-(i). But overall a clear increase in mastery of Importance Sampling over the previous year.